What Is Claimed Is:

1. A method for learning a generative model for text, comprising:

receiving a current model, which contains terminal nodes representing random variables for words and can contain cluster nodes representing clusters of conceptually related words;

wherein nodes in the current model are coupled together by weighted links, so that if a cluster node in the probabilistic model fires, a weighted link from the cluster node to another node causes the other node to fire with a probability proportionate to the link weight;

receiving a set of training documents, wherein each training document contains a set of words; and

applying the set of training documents to the current model to produce a new model.

2. The method of claim 1, wherein applying the set of training documents to the current model involves:

applying the set of training documents to the links defined in the current model to produce functions for weights for corresponding links in the new model; and

optimizing the functions to produce weights for links in the new model.

3. The method of claim 2, wherein for a given link, producing functions for a weight on the given link involves:

producing a function for the given link for each document in the set of training documents; and multiplying the functions for each document together to produce a function to be optimized for the given link.

4. The method of claim 3, wherein for the given link the function for a document is an approximation of the probability of the document's terminals firing as a function of the weight on the given link, keeping all other link weights in the model constant.

5. The method of claim 1, wherein the method further comprises iteratively:

considering the new model to be the current model; and

applying training documents to the current model to produce a subsequent new model.

6. The method of claim 5, wherein during an initial iteration, the method further comprises generating an initial current model from a set of words by:

generating a universal node that is always active;

generating terminal nodes representing words in the set of words; and

directly linking the universal node to the terminal nodes.

7. The method of claim 5, wherein each iteration uses twice as many training documents as the previous iteration until all available training documents are used.

8. The method of claim 1, wherein producing the new model additionally involves selectively introducing new links from clusters to nodes and from clusters to clusters.

9. The method of claim 8, wherein introducing a new link can involve:

considering a cluster that is likely to be active in generating a given document;

considering a new term in the given document, wherein the new term is not associated with the cluster; and

adding the new link between the cluster and the new term.

10. The method of claim 8, wherein introducing a new link can involve:

considering a first cluster that is likely to be active in generating a given document;

considering a second cluster that is likely to be active in generating the given document, wherein the second cluster is not associated with the first cluster; and

adding the new link between the first cluster and the second cluster.

11. The method of claim 1, wherein producing the new model additionally involves selectively introducing new cluster nodes into the current model.

12. The method of claim 11, wherein selectively introducing a new cluster node involves:

examining a given document;

creating the new cluster node;

creating links between the new cluster node and terminals in the given document; and

creating links between cluster nodes that are likely to have been involved in generating the given document and the new cluster node.

13. The method of claim 1, wherein producing the new model involves calculating an activation for each cluster node in each document, wherein the activation for a given cluster node indicates how many links are likely to fire from the given cluster node to other nodes.

14. The method of claim 1, wherein producing the new model involves renumbering clusters in the current model to produce a cluster numbering for the new model; and

wherein clusters that are likely to be active in generating more documents are assigned lower numbers that occur earlier in an identifier space, whereas clusters that are likely to be active in generating fewer documents are assigned higher numbers that occur later in the identifier space.

15. The method of claim 1, wherein applying a given document to the current model involves:

updating a summary variable for each cluster that is likely to be active in the given document, wherein the summary variable summarizes the probabilistic cost of the cluster linking to terminals not existing in the given document; and

for terminals that actually do exist in the given document, canceling the effects of corresponding updates to the summary variables.

16. The method of claim 1, wherein applying the set of training documents to the current model involves computing once for each cluster the probabilistic cost of the cluster existing in a document and triggering no words, and for each document applying this cost and subtracting the effects of words that do exist in the document.

17. The method of claim 1, wherein the probabilistic model includes a universal node that is always active and that has weighted links to terminal nodes and/or cluster nodes.

18. A computer-readable storage medium storing instructions that when executed by a computer cause the computer to perform a method for learning a generative model for text, the method comprising:

receiving a current model, which contains terminal nodes representing random variables for words and can contain cluster nodes representing clusters of conceptually related words;

wherein nodes in the current model are coupled together by weighted links, so that if a cluster node in the probabilistic model fires, a weighted link from the cluster node to another node causes the other node to fire with a probability proportionate to the link weight;

receiving a set of training documents, wherein each training document contains a set of words; and

applying the set of training documents to the current model to produce a new model.

19. The computer-readable storage medium of claim 18, wherein applying the set of training documents to the current model involves:

applying the set of training documents to the links defined in the current model to produce functions for weights for corresponding links in the new model; and

optimizing the functions to produce weights for links in the new model.

20. The computer-readable storage medium of claim 19, wherein for a given link, producing functions for a weight on the given link involves:

producing a function for the given link for each document in the set of training documents; and

multiplying the functions for each document together to produce a function to be optimized for the given link.

21. The computer-readable storage medium of claim 20, wherein for the given link the function for a document is an approximation of the probability of the document's terminals firing as a function of the weight on the given link, keeping all other link weights in the model constant.

22. The computer-readable storage medium of claim 18, wherein the method further comprises iteratively:

considering the new model to be the current model; and

applying training documents to the current model to produce a subsequent new model.

23. The computer-readable storage medium of claim 22, wherein during an initial iteration, the method further comprises generating an initial current model from a set of words by:

generating a universal node that is always active;

generating terminal nodes representing words in the set of words; and

directly linking the universal node to the terminal nodes.

24. The computer-readable storage medium of claim 22, wherein each iteration uses twice as many training documents as the previous iteration until all available training documents are used.

25. The computer-readable storage medium of claim 18, wherein producing the new model additionally involves selectively introducing new links from clusters to nodes and from clusters to clusters.

26. The computer-readable storage medium of claim 25, wherein introducing a new link can involve:

considering a cluster that is likely to be active in generating a given document;

considering a new term in the given document, wherein the new term is not associated with the cluster; and

adding the new link between the cluster and the new term.

27. The computer-readable storage medium of claim 25, wherein introducing a new link can involve:

considering a first cluster that is likely to be active in generating a given document;

considering a second cluster that is likely to be active in generating the given document, wherein the second cluster is not associated with the first cluster; and

adding the new link between the first cluster and the second cluster.

28. The computer-readable storage medium of claim 18, wherein producing the new model additionally involves selectively introducing new cluster nodes into the current model.

29. The computer-readable storage medium of claim 28, wherein selectively introducing a new cluster node involves:

examining a given document;

creating the new cluster node;

creating links between the new cluster node and terminals in the given document; and

creating links between cluster nodes that are likely to have been involved in generating the given document and the new cluster node.

30. The computer-readable storage medium of claim 18, wherein producing the new model involves calculating an activation for each cluster node in each document, wherein the activation for a given cluster node indicates how many links are likely to fire from the given cluster node to other nodes.

31. The computer-readable storage medium of claim 18, wherein producing the new model involves renumbering clusters in the current model to produce a cluster numbering for the new model; and

wherein clusters that are likely to be active in generating more documents are assigned lower numbers that occur earlier in an identifier space, whereas clusters that are likely to be active in generating fewer documents are assigned higher numbers that occur later in the identifier space.

32. The computer-readable storage medium of claim 18, wherein applying a given document to the current model involves:

updating a summary variable for each cluster that is likely to be active in the given document, wherein the summary variable summarizes the probabilistic cost of the cluster linking to terminals not existing in the given document; and

for terminals that actually do exist in the given document, canceling the effects of corresponding updates to the summary variables.

33. The computer-readable storage medium of claim 18, wherein applying the set of training documents to the current model involves computing once for each cluster the probabilistic cost of the cluster existing in a document and triggering no words, and for each document applying this cost and subtracting the effects of words that do exist in the document.

34. The computer-readable storage medium of claim 18, wherein the probabilistic model includes a universal node that is always active and that has weighted links to terminal nodes and/or cluster nodes.

35. An apparatus that learns a generative model for text, comprising:

a receiving mechanism configured to receive a current model, which contains terminal nodes representing random variables for words and can contain cluster nodes representing clusters of conceptually related words;

wherein nodes in the current model are coupled together by weighted links, so that if a cluster node in the probabilistic model fires, a weighted link from the cluster node to another node causes the other node to fire with a probability proportionate to the link weight;

wherein the receiving mechanism is configured to receive a set of training documents, wherein each training document contains a set of words; and

a training mechanism configured to apply the set of training documents to the current model to produce a new model.